Multivariate Behavioral Research

ISSN: 0027-3171 (Print) 1532-7906 (Online) Journal homepage: https://www.tandfonline.com/loi/hmbr20

aPRIDIT Unsupervised Classification with Asymmetric Valuation of Variable Discriminatory Worth

Linda L. Golden, Patrick L. Brockett, Montserrat Guillén & Danae Manika

To cite this article: Linda L. Golden, Patrick L. Brockett, Montserrat Guillén & Danae Manika (2019): aPRIDIT Unsupervised Classification with Asymmetric Valuation of Variable Discriminatory Worth, Multivariate Behavioral Research, DOI: 10.1080/00273171.2019.1665979

To link to this article: https://doi.org/10.1080/00273171.2019.1665979

Published online: 27 Sep 2019.

MULTIVARIATE BEHAVIORAL RESEARCH
https://doi.org/10.1080/00273171.2019.1665979

aPRIDIT Unsupervised Classification with Asymmetric Valuation of Variable Discriminatory Worth

Linda L. Golden[a], Patrick L. Brockett[b], Montserrat Guillén[c], and Danae Manika[d]

[a]Department of Marketing and of Business, Society and Government, McCombs School of Business, University of Texas at Austin, USA; [b]Department of Information, Risk and Operations Management, McCombs School of Business, University of Texas at Austin, USA; [c]Director of the Research Group on Risk in Insurance and Finance, University of Barcelona Av. Diagonal, Spain; [d]Newcastle University Business School, Newcastle upon Tyne, England

ABSTRACT

Sometimes one needs to classify individuals into groups, but there is no available grouping information due to social desirability bias in reporting behavior like unethical or dishonest intentions or unlawful actions. Assessing hard-to-detect behaviors is useful; however it is methodologically difficult because people are unlikely to self-disclose bad actions. This paper presents an unsupervised classification methodology utilizing ordinal categorical predictor variables. It allows for classification, individual respondent ranking, and grouping without access to a dependent group indicator variable. The methodology also measures predictor variable worth (for determining target behavior group membership) at a predictor variable category-by-category level, so different variable response categories can contain different amounts of information about classification. It is asymmetric in that a "0" on a binary predictor does not have a similar impact toward signaling "membership in the target group" as a "1" has for signaling "membership in the non-target group." The methodology is illustrated by identifying Spanish consumers filing fraudulent insurance claims. A second illustration classifies Portuguese high school student's propensity to alcohol abuse. Results show the methodology is useful when it is difficult to get dependent variable information, and is useful for deciding which predictor variables and categorical response options are most important.

KEYWORDS

Detecting hidden behavior; classification into non-self-disclosed behavior groups; unsupervised learning; non-parametric classification; asymmetric measures

Introduction

Sometimes people and institutions do not tell the truth. Individuals may lie or engage in fraudulent and/or illegal behavior (or generally may not truthfully acknowledge socially inappropriate behavior). Companies may also not willingly reveal undesirable or illegal behavior (e.g. bait and switch, false advertising, money laundering, etc.). People in socially unacceptable, potentially criminal or covert activity against social or legal norms will generally not disclose it willingly, so building statistical models to identify individuals engaging in "hidden behaviors" are difficult.[1]

It is useful, however, in many social science research environments to have a technique to identify individuals engaging in hard-to-detect behaviors, even while there may be no reliable set of data involving individuals identified as having engaged in this behavior, along with their corresponding covariates. There is also a societal interest in detecting illicit hidden behavior, as such behavior costs everyone money. For example, consumer fraud (an undisclosed hidden behavior) is hugely costly to society. A rough 2018 estimate by the Association of Certified Fraud Examiners (2019) is that companies worldwide lost

[1]Randomized survey response methods exist for eliciting self-identification of embarrassing or socially stigmatizing (or illegal) information by reducing social desirability bias (cf., De Jong et al., 2010), however, they do not work when people are likely to be dishonest in admitting socially undesirable actions like fraud even in a randomized setting. An approach that builds up an assessment of the likelihood of target behavior group membership based on exogenous indicators of membership is needed without assuming there is correct knowledge of an unavailable membership variable itself. Instead, observable predictor variables related to the target behavior can supply information about the actual behavior even when the dependent behavior variable itself is missing or untrustworthy.

CONTACT Patrick L. Brockett utpatrickbrockett@gmail.com Department of Information, Risk and Operations Management, McCombs School of Business, University of Texas at Austin, 2110 Speedway, Austin, TX 78705, USA.

close to $4 trillion to consumer fraud. A typical firm loses 5% of revenues (Association of Certified Fraud Examiners, 2019) to such non-self-disclosed behavior. Firms and governments work to identify (and perpetrators to hide) costly illicit behavior (from hackers to shoplifters to fraud), and social scientists work to identify socially undesirable or dangerous behavior that may be hidden. There are statistical techniques for identifying fraudulent behavior (Bolton & Hand, 2002), however for fitting the parameters of the statistical model used, these generally require information about subjects including both the predictor covariates and a target behavior membership indicator.

This article introduces an asymmetrical variable worth assessing classification technique (aPRIDIT) which extends PRIDIT (Principal Component Analysis of RIDITS, Brockett, Derrig, Golden, Levine, & Alpert, 2002) to help identify individuals who are engaged in problematic hidden social behaviors that will not be readily observed directly or admitted to. The technique introduced here does not require having a subset of identified records with known identified target behavior (i.e. it does not require a statistical training set with dependent and independent variables) upon which a "supervised training model" can be parametrically fit. It is a form of unsupervised learning for classification. The data used is a set of ordinal categorical predictor variables each of which has response categories ordered such that the lower the numbered response category, the stronger *a priori* is expected likelihood of the presence of the target behavior of interest.[2] All data in every study is collected for a purpose, and the data collected (and used) here is designed to give indications about who is (or is not) a member of the specified target behavior group of interest. Each input predictor variable is expected to be correlated with the target behavior we are attempting to uncover.

The methodology we present shows how to iteratively combine these predictor variables so as to obtain, for each individual, an overall "suspicion of target behavior group membership" score that linearly orders all respondents according to their suspiciousness level of having the behavior of interest (so classification of respondents can be achieved). The methodology also provides an overall assessment of each predictor variable's worth in discriminating between the respondents with the target behavior of interest and those without the target behavior of interest. Moreover, within each predictor variable, it identifies for each categorical response option of each categorical predictor variable how important selection of that particular categorical response option is for the classification. For example, for a binary 0/1 or yes/no-variable predictor, it may be that a "yes" is a better predictor of target group membership than a "no" on the same question is indicating non-target-group membership. Not all categories carry similar information about target group membership (or non-membership), and this technique delineates the relative importance of individual categories (not just the worth of the total variable) for discriminating among groups.

Identifying hidden behavior (a.k.a. classifying without dependent variable information)

Different data-availability structures contexts have different established statistical techniques that can be used to address classification problems. These techniques can be dichotomized into two categories. In the terminology of machine learning, these two techniques types are referred to as "supervised" and "unsupervised" learning methodologies. Supervised learning can occur when the data supplies a "training set" of available records that includes both dependent (target group membership identification) and independent (predictor) variable observations. This training set can be used, together with parametric model building, for classification wherein the parameters of the model are optimally fit to the known data set having both dependent and independent variables to improve classification accuracy. Examples of supervised classification techniques include neural networks, discriminant analysis, probit, logistic regression, support vector machines, and others.

Unsupervised learning classification algorithms, on the other hand, derive classification results using "unlabeled data" without having access to a known and verified dependent variable, and do not have full training sets with a targeted dependent variable available for parameter fit. Unsupervised classification techniques include cluster analysis, *k*-mean cluster analysis, Kohonen Self-Organizing Feature Maps (Kohonen, 1989), temporal or time series clustering (Gates, Lane, Varangis, Giovanello, & Guiskewicz,

[2]It is critical that the assumption holds that the response categories of the set of predictor variables are all ordered in terms of the a priori expectation that a lower category response is more indicative of target group membership. It is not required that all predictor variables have the same "worth" in discriminating between more suspicious and less suspicious behavior, but the wording of the variable should be constructed such that lower category responses are more suspicious. Essentially, in kind of an iterative "wisdom of the crowds" type statistical aggregation of the ensemble of the variable answers, when they are all ordered in the same way, can be used to extract classification without ever having observed an actual dependent variable (target behavior or no target behavior).

2017), and some others that group without access to a dependent variable.[3]

Here we introduce an unsupervised technique (aPRIDIT) that follows a line of development of unsupervised learning. The available data set contains discrete ordered categorical variables all of which, by construction of the predictor variable design, are expected to be related in the same ordered direction to the target behavior upon which the classification is to be performed (PRIDIT analysis for classification). We do not have an observation of the target behavior group classifier for the individuals in the data set (it is unsupervised). We also present additional new information about the relative classification value of each of the individual categorical responses within a predictor variable, and that individual category's worth for predicting target behavior group membership.

The RIDIT and PRIDIT numerically scoring methods for ordinal category variables

Non-numeric qualitative ordinal categorical variables (such as those used in many social science contexts) do not have apparent numbers associated with their categories. Numbers, however, are needed to render the variable capable of being used for subsequent numerical statistical analysis like *t* test, logistic regression, etc. In these situations, the analyst must choose a "numerical scoring system" and assign numbers to categories for many types of subsequent analyses. The RIDIT (*R*elative to an *I*dentified *D*istribution *T*ransformation) categorical data scoring technique was developed in the epidemiological literature for analysis of ordinal but non-metric qualitative data (like "degree of paleness") common in medical studies (Bross, 1958). Such ordinal but non-metric qualitative data are also common in social sciences (e.g. Likert scale categorical variables). Instead of labeling individual categories with raw integer values like 1, 2, 3, … as is often done, Bross developed so-called RIDIT scoring. It is a data-driven transformation of the category probability values developed to provide an increasing numerical score value for each category that better allows for subsequent standard statistical analysis (like *t* tests) to be performed. As Bechtel (1985) notes, the alternative commonly used raw integer scoring method forces a questionable tacit assumption of equal subjective distance between the adjacent response categories in the analysis. RIDIT scoring avoids this assumption. The RIDIT score Bross assigned for categorical response option k on a variable V is equal to the observed proportion of respondents in all categories below k plus one-half the proportion of respondents in category k. It is related to performing a relative rank transformation for the categorical variable (see Golden & Brockett, 1987).

PRIDIT scores used in this current paper are a variant of RIDIT scores, and are developed by normalizing (via linear transformation) the RIDIT score in such a manner that the PRIDIT score is always bounded between -1 and $+1$ and has a mean value of 0.[4] This zero-mean centering property allows comparability across variables when we combine them subsequently. If the tth predictor variable V_t has k_t possible response categories and the empirical proportion of data falling into each of the response categories is $(\hat{p}_{t1}, \hat{p}_{t2}, \ldots, \hat{p}_{tk_t})$, then the numerical PRIDIT score (designated by B_{tk}) assigned to category k based on the data is

$$B_{tk} = \sum_{j<k} \hat{p}_{tj} - \sum_{j>k} \hat{p}_{tj} \quad (1)$$

Thus, if a respondent answers (or falls into), the kth response category on predictor variable V_t, the numerical value assigned to this respondent's answer is not the integer score "k" as raw integer numerical scoring would give it, but rather is B_{tk}, the empirical proportion of respondents overall who, on variable V_t, responded in a category of lower ordinal category than k, minus the empirical proportion of respondents overall who responded in a category of higher ordinal category than k. This is the verbal rendition of Equation (1), and it gives a measure of how relatively "extreme" a response category is. A positive value for B_{tk} means more respondents gave responses in a lower category than gave responses in a higher category than k. The closer B_{tk} is to 1.0, the more extreme the response is in that almost everyone responds in lower ordinal category than k. If you respond in category k when B_{tk} is close to 1.0, you are anomalous in that your response category is ordinally high relative to what is normal. Likewise, the closer B_{tk} is to -1.0, the

[3]The unsupervised label refers to the analysis method and not the composition of data set itself. All analysts select variables for their data set pertinent to the task being performed. The dataset is then analyzed by the chosen technique, which can be supervised in nature or unsupervised, depending on the existence or non-existence of a dependent variable.

[4]The name PRIDIT scoring comes from the use of this scoring method together with a Principal Components Analysis to develop an unsupervised learning method for classification (see the subsequent theorem for where principal components analysis arises) when the dependent variable is not known (Brockett et al., 2002). PRIDIT analysis is discussed subsequently. PRIDIT scores were derived because the PRIDIT analysis depends on the zero- mean centering of variables for consistent treatment among variables. This is why we transformed RIDIT scores rather than using them directly.

more extreme the response is in that almost everyone responds in higher ordinal category than k. If you respond in category k when B_{tk} is close to -1.0, you are also anomalous in that your response category is ordinally low relative to what is normal. Integer scores for ordinal categories does not have this "relative extremeness" interpretation (cf. Brockett, 1981; Brockett & Golden, 1992).

A numerical example can illustrate the implementation of (1). Suppose variable 1 is a 5-point Likert scaled predictor variable with categories from extremely unlikely to extremely likely. Suppose further that the empirical responses probabilities to the five categories are (0.20, 25, 0.30, 0.20, 0.05). Then from (1) the PRIDIT scores for the five categories are $B_{11}=0-(0.25+0.3+0.2+0.05)=-0.80$, $B_{12}=0.20-(0.3+0.2+0.05)=-35$, $B_{13}=(0.20+0.25)-(2.0+0.05)=+0.20$, $B_{14}=(0.20+0.25+0.3)-0.05=+0.70$, and $B_{15}=(0.20+0.25+0.30+0.20)-0=+0.95$. Accordingly, for variable 1, instead of assigning the raw integer scoring to the ordinal categories as (1, 2, 3, 4, 5), we assign the PRIDIT scores (-0.8, -0.35, $+0.2$, $+0.7$, $+0.95$). The relationship between the RIDIT score R_k for category k and the PRIDIT score B_k for category k is $B_k = 2\ R_k - 1$.

The PRIDIT scoring method has several desirable properties, in addition to quantifying how relatively extreme a given response is. For example, it can be calculated that the empirical distribution for each scored variable V_t is "centered" in that it has mean value $\sum_k B_{tk}\hat{p}_{tk}=0$.

In the applications here, the ordinal, possibly qualitative, categorical variables collected for study inclusion are all expected to be indicators of the likelihood or suspiciousness of a certain target behavior that is desired to be identified, but which cannot be directly observed. All variables have the directionality of the categories ordered such that the lower the response category number, the higher the suspicion that an individual respondent belonging in that category exhibits the target behavior of interest. Membership in the target group that exhibits this behavior becomes increasingly less likely as the respondent responds in increasingly higher categorical response categories. Intuitively, the PRIDIT score assigns a numerical value to a categorical response option for the predictor variable that is a reflection of how relatively extreme a suspiciousness level is attached to a respondent who falls into that category relative to other respondents. A negative PRIDIT score number for category k means more respondents are in higher (less suspicious) response categories than are in lower (more suspicious) response categories than k. Put otherwise, since the categories are ordered monotonically decreasing with respect to suspiciousness likelihood (of target group membership), if more respondents are in higher categories than in lower categories, then the respondent who answered k is relatively more suspicious of belonging to the target suspiciousness group on that variable than are other respondents. The larger the absolute value of a PRIDIT score, the more extreme (relative to others) it is to belong to that category. Large negative values provide a stronger indication (suspicion of target group membership) relative to other respondents. In our numerical example, $B_{11}=-0.80$ so a category 1 response is very suspicious of the respondent exhibiting the target behavior (since 80% of respondents respond in a less suspicious category). On the other hand, $B_{15}=+0.95$ indicating a category 5 response is not very suspicious of target behavior group membership (since 95% of respondents respond in more suspicious categories than the respondent did).

It can be proven mathematically (Golden & Brockett, 1987) that this PRIDIT scoring method provides a consistent increasing numerical scoring method that yields an empirical distribution closest (in a Kolmogorov-Smirnoff distance sense) to a uniform distribution over $[-1,1]$ while maintaining a zero expected value. Also, PRIDIT scoring can be characterized mathematically as the unique way to assign numerical scores to rank ordered categories that satisfy certain intuitively reasonable assumptions that a numerical scoring mechanism for an ordinal categorical variable should satisfy (Brockett, 1981; Brockett & Levine, 1977). Moreover, the performance of PRIDIT scoring when using normality-based statistics, such as regression, is better than other numerical scoring methods examined (cf., Brockett & Golden, 1992; Golden & Brockett, 1987). Also, the Kolmogorov-Smirnoff distance to the multivariate normal distribution is smaller in a real-dataset of rank ordered categorical variables when they are scored with PRIDIT than when scored with other commonly used methods of assigning ordinal scores to categories prior to analysis.

The PRIDIT and aPRIDIT classification methodology and an asymmetric aPRIDIT measure of individual response category discriminatory value assessment

Categorical scoring algorithms described previously are one thing, and unsupervised classification

algorithms we describe next are another. These two conceptual developments are related in our derivation since we must first, as a data quality check starting point, verify that the response categories used in all the predictor variables are ordered according to *a priori* assessment of the underlying variable of interest. That is, the response categories are such that a lower number categorical response is expected to be more likely to arouse suspicion of the respondent having target group membership. PRIDIT *scoring* is done on all predictor variables. Although the classification algorithm presented here does not require knowledge of an observed dependent variable indicating target group membership (i.e. it is unsupervised classification), it absolutely must have carefully selected predictor variables that are a priori selected to be indicative of that group membership as a substitute. Furthermore, these predictor variables should be constructed with a great deal of thought based on subject area expertise. It is this expertise utilized in constructing consistent predictor variables that stands in lieu of (in the procedure we outline) the knowledge of group membership utilized in supervised methods.

By adding together the individual PRIDIT scores of a respondent for all of the predictor variables, we can order the respondents according to their overall "suspiciousness of target group membership" (Brockett et al., 2002). Intuitively this is akin to the process a teacher might use when summarizing student scores over multiple tests which all measure different aspects of the same latent dimension (knowledge of the material) so as to classify students into groups. While this is useful for classification, we can do better by iteratively updating this process in steps. This is the essence of the classification methodology introduced here, and these steps are presented next.

Steps in aPRIDIT analysis[5]

The first step in the analysis is to calculate a PRIDIT score for each of the response categories for each of the predictor variables V_t. These PRIDIT scores are the numerical values given previously by (1). The PRIDIT score for a categorical response option k in a predictor variable V_t measures how extreme a response answer k is. If B_{tk} is relatively large negatively, then most individuals answer in a higher or "not as suspicious of membership in the target behavior group" as a category k respondent. In such a situation a person answering in the kth category gets much more weight as an outlier in terms of their "category k" answer indicating "higher suspicious of membership in the focal target behavior group." Similarly, if B_{tk} is a relatively large positive number this indicates the extremeness of a response in category k for indicating non-target behavior group membership (most respondents are more suspicious on this variable than the particular respondent who answered k). An analogy can be made to the process used in "anomaly detection" in data mining (Chandola, Banerjee, & Kumar, 2009). The PRIDIT score gives a measure of how anomalous a response is, and since the variables are oriented inversely toward suspiciousness, of how extremely suspicious the respondent's response is.

Step 1: Assign PRIDIT numerical scores to predictor variable categories

The PRIDIT scoring procedure (1) transforms the categories into numerical values that reflect the relative "abnormality" of particular response categories among sets of individual respondents. The more "unusual" a response (as reflected by the absolute value of the size of the PRIDIT score), the more information is contained in the response. PRIDIT scoring avoids assigning integer values to categories in an *ad hoc* fashion and it improves the statistical characteristics of the scored data (see Brockett & Golden, 1992; Golden & Brockett, 1987).

Step 2: For each entity obtain an initial overall relative score position along the latent suspiciousness dimension

Let $\mathbf{F} = (f_{it})$ denote the $N \times m$ matrix of individual PRIDIT variable scores for m predictor variables V_t, $t = 1, 2, \ldots, m$, and N individuals or respondents $i = 1, 2, \ldots, N$. Rows delineate respondents, columns delineate predictor variables, and the entry $f_{it} = B_{tk}$ if respondent i has exhibited categorical response level k on variable V_t. All analysis uses PRIDIT category scores instead of integer scores for the matrix $\mathbf{F}$.

The iterative classification methodology process begins as follows. For each individual an initial overall target group membership suspicion score is obtained by adding the individual's predictor variable PRIDIT scores. (This is akin to simply adding together the grades on a series of exams to get an overall grade assessment). In matrix notation, let $\mathbf{W}^{(0)} = (1,1, \ldots, 1)'$ be a $m \times 1$ vector. The initial vector of overall summative scores for the N entities or individual respondents across the m predictor

[5]The first seven steps in the analysis are the same for PRIDIT and aPRIDIT, so the whole process will be described for aPRIDIT alone. Neither PRIDIT analysis nor aPRIDIT analysis has appeared in the social science literature, so it is fully detailed here.

variables is $\mathbf{S}^{(0)}=\mathbf{FW}^{(0)}$. This gives a numerical overall suspiciousness score for each respondent along the latent dimension of target group membership suspicion. These overall scores are deemed proxy aggregate assessments of "suspiciousness of target behavior" since, by construction using area expertise, each of the individual predictor variable was categorically ordered to have lower response categories be more indicative of target behavior group membership.

Step 3: Obtain an initial assessment of predictor variables' discriminatory ability along the latent suspiciousness dimension

Taking the normalized scalar product of the individuals' *overall* summative latent dependent variable scores $\mathbf{S}^{(0)}$ with their *individual variable* V_t *scores* provides a consistency measure of the individuals' responses on the predictor variable V_t *vis a vis* their position on overall latent dimension score. This is similar to Cronbach's measure of reliability in questionnaire analysis for assessing individual question consistency with overall questionnaire scores (cf., Gliem & Gliem, 2003). Conceptually it is like correlating individuals' predictor variable scores with their overall summative scores, the correlation being taken across the N individuals. In matrix notation $\mathbf{W}^{(1)}=\mathbf{F}'\mathbf{S}^{(0)}/||\mathbf{F}'\mathbf{S}^{(0)}||$. This normalized scalar product provides a system of relative "weights" for predictor variables, where the components of $\mathbf{W}^{(1)}$ give the normalized product of the predictor variable V_t with overall the total latent suspiciousness of target group membership score. The weights can be interpreted as a measure of the "worth" of the predictor variable for overall ranking of respondents. Similar to correlation, the variable weights $\mathbf{W}^{(1)}$ measure the consistency of scores on individual variables with the overall summative scores. For any t, a larger value for $W_t^{(1)}$ indicates higher consistency of V_t values with overall scores. This offers an assessment of variables' discriminatory ability roughly akin to item analysis done on multi-question examinations to assess the worth of individual questions in a test: Do high test scoring individuals score high on question V_t and low scoring individuals score low on question V_t? If so, V_t is a worthwhile test variable.

Step 4: Revise the overall relative individuals' scores along the latent dimension to reflect the differential discriminatory ability of the predictor variables

Good and bad predictor variables should not get equal weights in forming an overall suspiciousness score for the individual respondent. Since we now have a numerical assessment $W_t^{(1)}$ of which predictor variables provide stronger signals of overall consistency with the latent suspiciousness dimension, it makes sense to use these new variable weights $W_t^{(1)}$ to compute a revised overall weighted suspiciousness score for each respondent. Consequently, we compute a new weighted summative score for each respondent using the new weights $W_t^{(1)}$. This gives more importance or weight to "better discriminating" variables. In matrix notation, the new weighted score vector for entities or individual respondents obtained by using newly calculated weights $\mathbf{W}^{(1)}$ is $\mathbf{S}^{(1)} = \mathbf{FW}^{(1)}$.

Step 5: Obtain a revised assessment of predictor variables' latent dimension discriminatory ability using new revised respondent overall weighted scores

Using Step 4's "refined" score, $\mathbf{S}^{(1)}$, for each entity's relative position along the latent suspiciousness dimension, we can obtain an even better numerical assessment of each of the predictor variables relative ability to discriminate along the latent variable dimension. We employ the same process as in Step 4, namely "correlating" the predictor variable scores of individuals with their new weighted summative overall scores $\mathbf{S}^{(1)}$. Mathematically, this yields a new vector of weights, $\mathbf{W}^{(2)}=\mathbf{F}'\mathbf{S}^{(1)}/||\mathbf{F}'\mathbf{S}^{(1)}||$.

Step 6: Iteratively revise the entities' overall relative scores and predictor variables' latent dimension discriminatory ability scores

The new predictor variable weights $\mathbf{W}^{(2)}$ obtained in Stage 5 are used in Stage 6 to get yet another revised weighted overall entity score vector $\mathbf{S}^{(2)}=\mathbf{FW}^{(2)}$. This new score vector $\mathbf{S}^{(2)}$ in turn can again be correlated with the vector of individual predictor variables scores to get revised vector of weights $\mathbf{W}^{(3)}=\mathbf{F}'\mathbf{S}^{(2)}/||\mathbf{F}'\mathbf{S}^{(2)}||$. Moreover, this process can continue. The revised relative predictor variable weight vector at stage $n+1$ is derived from overall entity scores at iterative stage n, via $\mathbf{W}^{(n+1)}=\mathbf{F}'\mathbf{S}^{(n)}/||\mathbf{F}'\mathbf{S}^{(n)}||$, and the revised overall weighted summative entity score vector at stage $n+1$ is developed from weights for the predictors V_t at stage $n+1$ via $\mathbf{S}^{(n+1)}=\mathbf{FW}^{(n+1)}$.

Theorem. The sequence of predictor variable weights $\{\mathbf{W}^{(n)}\}$ converges to a final weight vector $\mathbf{W}^{(\infty)}$ and the sequence $\{\mathbf{S}^{(n)}\}$ of overall weighted summative individual respondents' suspiciousness scores converges to a limiting score vector $\mathbf{S}^{(\infty)}$. Moreover, the final weight vector $\mathbf{W}^{(\infty)}$ is the first principal component of the PRIDIT scored matrix $\mathbf{F}'\mathbf{F}$.

Proof. By backward mathematical induction we calculate $\mathbf{W}^{(n)}=(\mathbf{F}'\mathbf{F})^{\mathbf{n}}\mathbf{W}^{(0)}/||(\mathbf{F}'\mathbf{F})^{\mathbf{n}}\mathbf{W}^{(0)}||$ where $\mathbf{F}'\mathbf{F}$ is a symmetric matrix. As a normalized sequence of increasing powers of a symmetric matrix, this iterative

process converges to a limit and will yield a single unique final set of predictor variable weights $\mathbf{W}^{(\infty)}$ (see Golub & Van Loan, 1996 section 7.3 for a proof of convergence). Moreover, the limiting vector of this normalized sequence of increasing powers of a symmetric matrix is the first eigenvector (principal component) of that symmetric matrix (cf., Golub & Van Loan, 1996), and hence $\mathbf{W}^{(\infty)}$ is a first principal component of $\mathbf{F}'\mathbf{F}$. Since $\mathbf{S}^{(n)}=\mathbf{F}\mathbf{W}^{(n)}$, and $\mathbf{W}^{(n)}$ converges, the final overall weighted summative score vector $\mathbf{S}^{(n)}$ also converges, and $\mathbf{S}^{(\infty)}=\mathbf{F}\mathbf{W}^{(\infty)}$. In practice, following the results of the Theorem, a direct application of a Principal Components Analysis (PCA) computer program[6] can be used instead of iteration. We choose to use a PCA computer package with the Varimax rotation for determining the values for $\mathbf{S}^{(\infty)}$ and $\mathbf{W}^{(\infty)}$. Note that $\mathbf{W}^{(\infty)}$obtained from the Varimax rotation first principal component analysis is not unique in algebraic sign (i.e. if $\mathbf{W}^{(\infty)}$ is a Varimax first principal component, then the negative $-\mathbf{W}^{(\infty)}$ is also a Varimax principal component). Accordingly, we take the positive Varimax first principal component given by the computer program making sure that it has primarily positive values by, if necessary, multiplying the weight vector obtained from the computer program by -1 in order to get positive weights.[7]

Note that the fact that the weights (variables' worth for assessing suspiciousness of target behavior group membership) obtained by the intuitive iterative process is the principal component of the PRIDIT matrix is why the name PRIDIT (Principal Component Analysis of RIDITS) was used. While we did not start by using principal components to determine weight (rather we started from the intuitively appealing iterative updating), we arrived at weights that are principal components. Using principal component analysis to "weight" variables is not new (Daultry, 1976). However, in general, there is no guarantee that principal components weights will have meaningful statistical interpretations. By using the PRIDIT scoring system it can be proven (in conjunction with a local independence assumption from latent trait analysis literature) that there is a useful interpretation for $W_t^{(\infty)}$ in terms of discriminatory power of the variable V_t (see Brockett et al., 2002 for details).

Step 7: Obtain entity final weighted suspiciousness score and use final respondent scores for group classification.

Steps 1–6 create an empirical aPRIDIT (or PRIDIT) relative suspiciousness score vector $\mathbf{S}^{(\infty)}$. Respondents can be rank ordered by their relative position along the latent dimension of interest (suspiciousness of target group membership) according to their scores $\mathbf{S}^{(\infty)}=\mathbf{F}\mathbf{W}^{(\infty)}$. Importantly, the suspiciousness score obtained for each respondent is metric level data and hence can be used in conjunction with other outside information, such as demographics, for further metric level statistical analysis if desired. Other unsupervised classification methods such as cluster analysis and Kohonen's Self-Organizing Feature Map do not produce metric level data sores for individuals and do not have this desirable characteristic.

For classification of respondents, we need to go further than simply linearly ordering (although this is important in its own right). The process for dichotomously classifying respondents into the target or non-target group of interest depends upon whether or not we know the expected proportion, θ, of entities in the sample from the lower ranked target behavior group.

If the proportion θ is known (e.g. we might know or expect from previous research what proportion of population members are in the target group), then respondents can be ordered by their overall suspicion score $S^{(\infty)}$ and the first θN entities are assigned to the lower group (the target group of interest) and the remainder into the non-target group. This is akin to using an informative Bayesian prior. It is also analogous to the "priors proportional" versus "priors equal" supervised learning choice in logistic regression. Methods for estimating θ from the data are available in AI, Brockett, Golden, & Guillen (2013).

If θ is unknown, we simply divide the entities into groups according to whether they have positive or

[6]When implementing the PCA computer program we choose the Varimax rotation since it is the orthogonal rotation that provides the first principle component (eigenvector) which maximally explains the overall variability. Since by construction the predictor variables all have the same (ostensibly primarily unidimensional) target latent dimensional ordering of responses from low suspiciousness to high suspiciousness of target behavior, this is the first principal direction, we want the weights (PCA eigenvector) that maximally explain variability in this data, which was designed primarily along the single "suspiciousness of target group behavior" dimension. This is the Varimax rotation.

[7]It is also possible that some isolated components of $\mathbf{W}^{(\infty)}$ could still be negative even while the vast majority of the components of $\mathbf{W}^{(\infty)}$ are positive. This can occur if the experts constructing the predictor variables expected a positive relationship with the target behavior, but they were wrong and, in fact, the relationship was negative. It is worth noting, however, that for obtaining rank ordering of the individuals' scores, aPRIDIT is self-correcting in that if the expected suspiciousness order coding were incorrect then the weight for this predictor variable will come out negative. Upon multiplying the predictor weight by the score, the contribution to the overall score becomes correctly signed and adds correctly to the individuals' overall aPRIDIT score. An easy solution for variables that were incorrectly ordered is to reverse the role of "0" and "1" in the analysis (so, for example if a "0 = yes" resulted in a negative weight, then use "0 = no" and the predictor variable weight changes to positive resulting in a correctly signed contribution to the overall respondent score used for classification.

negative overall summative weighted $\mathbf{S}^{(\infty)}$ score. The rationale for this is that variables' ordinal response categories were initially designed to have the left-most category corresponding to more expected suspicion of the respondent exhibiting the hidden target behavior. A negative value for B_{tk} indicates there are more respondents with less suspiciousness of target group membership than there are with more suspiciousness, so if an overall entity's weighted summative score is negative, this indicates the evidence is more likely than not to be consistent with target group membership. Such a respondent with a negative weighted summative score is then classified as belonging to the target group. This classification rule using a negative/positive dichotomy is akin to a non-informative or vague prior in Bayesian analysis.

Step 8: aPRIDIT asymmetrical assessment of predictor variable response categories' impact for classification

Because predictor variables were scaled so *a priori* a lower category response was expected to be more indicative of the respondent belonging to the target group of interest, a negative PRIDIT score B_{tk} on the variable V_t increases the overall likelihood that the respondent's total weighted suspicion score $\mathbf{S}^{(\infty)}$ will be sufficiently negative that the entity will be labeled as belonging to the target group.

For any respondent, the size of the absolute value of their B_t value recorded for V_t indicates how divergent their response is from the norm of the other respondents on this predictor V_t. Being divergent on an unimportant (for discriminating) variable, however, is less informative for classification purposes than being divergent on an important discriminatory variable. We use the predictor weight $\mathbf{W}^{(\infty)}$ to provide an assessment of the relative *overall importance* of the individual indicator V_t and the PRIDIT score for the category to provide an assessment of how extreme the particular categorical response is from the norm. Our asymmetric measure assesses how much of this overall suspiciousness score total for an entity is contributed by a particular response to a particular predictor variable, a category by category assessment for each predictor variable.

The measure of the *asymmetric* contribution to classification made by a particular categorical response k on variable V_t. is obtained by multiplying the extremeness of the response category (measured by B_{tk}) by the importance of the variable for discriminating $W_t^{(\infty)}$. For a predictor variable V_t, and for its "category k" response option, we calculate the value $B_{tk}\times W_t^{(\infty)}$. This measures the relative importance of this particular category for its contribution to overall classification.

We focus first on the various categorical responses that made contributions toward an overall target group membership assessment (those categories with negative values of $B_{tk}\times W_t^{(\infty)}$ since these are the relatively more suspicious responses). The relative importance of a particular "category k" response on V_t which has a negative value for B_{tk} is ranked relative to all other variables' categorical responses on variables V_s that have negative B_s values (and which therefore have also made negative contributions to the overall score). The relative importance of the category k response on variable V_t is assessed by the *rank* (across predictor variables) of $B_{tk}\times W_t^{(\infty)}$ among all the negative values of $B_{sj}\times W_s^{(\infty)}$ (which are the possible contributors to increased suspiciousness of the individual belonging to the target behavior group). Some categories in some variables are very important for a target group suspiciousness classification, and this methodology points out the relative importance of individual response categories in this target group classification.

Similarly, some responses will be strong indications of non-target group membership. We determine these categories in an analogous manner by ranking the value of $B_{tk}\times W_t^{(\infty)}$ among all the positive values of $B_{sj}\times W_s^{(\infty)}$ (which are the contributors to positive overall scores supporting the overall classification of non-target group).

Summarizing, the rank of the negative value $B_{tk}\times W_t^{(\infty)}$ for a "category k" answer on variable V_t among all negative values of $B_{sj}\times W_s^{(\infty)}$ tells how relatively important this "category k" answer on V_t was to the overall negative score contributions for the individual. For a negative PRIDIT scored category k, this rank measure incorporates both the degree of extremeness B_{tk} of the particular response on V_t and the importance $W_t^{(\infty)}$ of the predictor variable V_t for predicting target group suspiciousness level. The rank measure asymmetrically slices the predictor variable value $W_t^{(\infty)}$ into components of categorical response level contributions to classification results.[8] These will be illustrated with real examples subsequently.

A computer code in R is available from the authors to perform this analysis.

Further discussion of the value of asymmetric assessment of variable category worth

Both PRIDIT and aPRIDIT give the same classification rule. The additional contribution of aPRIDIT

[8]aPRIDIT analysis also responses to Woodside's (2013) call for more asymmetric measures in data analysis.

over PRIDIT is in the development of an asymmetric measure of individual category worth for predicting target group membership (and a refinement in the interpretation of the value of a predictor variable). Hence, this is worth additional discussion. The original PRIDIT measure $W^{(\infty)}$ of the worth of a variable for predicting group membership, like other statistical methods based on categorical ordinal predictor variables, is a single overall measure of a variable's worth for discriminating between groups.

The distinguishing feature of aPRIDIT for assessing variable worth is the recognition that even within the categories of the same predictor variable; different predictor variable response categories contain different amounts of information about classification group membership. When only a single average measure of predictor variable worth is presented, as in original PRIDIT or in standard covariance based statistical predictor variable worth assessments that correlate the variable with the overall score, this information is blended together (and perhaps obfuscated).

We clarify this issue by example. For this, we look at data from the USA Massachusetts Automobile Insurance Fraud Bureau data discussed in Brockett et al. (2002). The goal was to detect individuals filing false bodily injury claims for automobile insurance, and to find which predictor variables were best at identifying fraudsters. Among others, three variables gathered were "Insured has history of prior claims" (Yes/No), "Claimant had old low value vehicle" (Yes/No) and "No police report at scene" (Yes/No). In each case, the first response category is expected to indicate a higher likelihood or degree of fraud suspicion. While each of these predictor variables had high blended value overall (high $\mathbf{W}^{(\infty)}$) for predicting suspiciousness of fraud, the discriminatory information value within the individual categories is not similar. A "No" (high category) response on prior claim history for signaling that the customer is *not* committing fraud is not as strong a signal for non-fraud as is a "Yes" on this same variable for signaling fraud. Affirmation of prior claims contains more information about fraud than non-affirmation contains about non-fraud. By focusing solely on the aggregated overall (symmetric averaged) measure of predictor variable worth, the information is lost that just one category of that variable accounts for variable's usefulness in predicting suspiciousness level. There is asymmetry between the value of a "Yes" category signaling target class membership versus the value of a "No" category on the same predictor for signaling non-target-class membership. Likewise, by performing aPRIDIT analysis we find that a "No" on "Claimant had old low value vehicle" is a much stronger indicator about the claimant *not* engaging in fraud than "Yes" on this same variable is as an indicator about the claimant affirmatively engaging in fraud. Essentially, *having* an old low value vehicle involved in the accident provides little information supporting fraud suspicion (as both fraudsters and non-fraudsters often have old low value vehicles). However, having a newer high valued car *does* provide much more information that fraud is *not* involved since non-fraudsters are more likely to have a substantial asset involved in an accident than are fraudsters. There is little fraud/non-fraud classification information in having an old vehicle in the accident, but a lot of information in having a newer vehicle (the new vehicle predicts and the old does not). Finally, not calling the police at the time of the accident is not very suggestive of fraud (many accidents go unreported) whereas calling the police to the accident *is* suggestive of *non-fraud (*fraudsters do not want the police there).

These examples illustrate how the particular response variable category into which the respondent belongs can provide important additional classification information beyond the importance weights $\mathbf{W}^{(\infty)}$ themselves that could be lost when aggregated across the response categories so as to obtain a single overall predictor variable assessment of worth. There is non-symmetry in the discriminatory value of the different categorical responses.

Defining and measuring this asymmetric category-by-category variable worth value is a motivating factor for developing an asymmetric assessment of suspiciousness (the contribution aPRIDIT makes beyond PRIDIT). The name aPRIDIT was chosen to emphasize that it provides an asymmetric and individual response category dependent measure of predictor variable worth. aPRIDIT goes a step further than PRIDIT since it not only gives the same unsupervised group membership assessment as PRIDIT (i.e. it is unsupervised classification) but it also provides statistically asymmetric information on a predictor variable's individual response category value for reaching the classification determination.

Two examples illustrating the use of the aPRIDIT methodology

Since PRIDIT has not been in the social science literature previously and aPRIDIT gives an extension of PRIDIT, for better understanding we present two studies to illustrate the computation and

interpretation of PRIDIT classification and asymmetric variable category classification worth assessment. Our illustrations use two international databases related to issues having social science importance and which involve identifying hard-to-detect behaviors: individuals defrauding an insurance company (in Spain) and teenagers with excessive drinking (in Portugal). In both settings, a reliable dependent variable classification is lacking or questionable due to legal or social desirability bias.

Study one: detecting consumers defrauding a Spanish insurer

The first illustration uses aPRIDIT to classify individuals filing potentially fraudulent automobile accident damage claims based on their fraud suspiciousness predictor variables. The data is a proprietary sample of 1995 automobile insurance claims filed between 1993 and 1996 by customers of a large Spanish insurer (cf., Artís, Ayuso, & Guillén, 1999, 2002). The predictor variables were restricted to those available either from the policy itself, or from the initial claim filing because such information is available early in the claims handling processes. Temporally, this is when signaling fraud suspiciousness for decision-making related to the existence/non-existence of claim fraud is most critical.

Claim files included 18 binary fraud indicators selected by company experts as indicative of heightened suspiciousness of fraud or because these experts expect this information to be relevant for identifying fraudulent claims. Binary predictors were selected because they could be easily and rapidly checked off by an assessor/adjuster. The categories for the fraud indicators were structured so that a low category response (a "0") was expected to raise more suspicion of fraud than a high category response (a "1"). Table 1 presents the variables; the calculation of the PRIDIT scores B_{t0} and B_{t1}, the calculation of predictor variable discrimination ability measure $W_t^{(\infty)}$, and the asymmetric rank measures of response category importance for a fraud determination.

Following the aPRIDIT methodology steps, all 1,995 claims were PRIDIT scored and ranked according to their aPRIDIT overall "fraud suspiciousness" score, and the relative worth of each individual fraud predictor variables was assessed via the weight vector $W_t^{(\infty)}$ which measures its ability to distinguish between highly suspicious and non-suspicious claimants. Table 1 presents the analysis ordered by the value of the predictive discrimination ability measure $W_t^{(\infty)}$ for each potential fraud assessor variable t.

For each predictor in column 1, column 2 gives the claims count (out of 1995 claim files) for which the lower ranked category (a 0) was checked off indicating higher suspiciousness of fraud. From Equation (1) the PRIDIT scores B_{t0} and B_{t1} for each response category are obtained. Columns 3 and 4, respectively, give the resultant PRIDIT scores for categories "0" and "1". For each variable, the discriminatory power measure $W_t^{(\infty)}$, shown in column 5, was calculated by taking the Varimax rotation first principal component of the $\mathbf{F}'\mathbf{F}$ matrix (analogous to iteratively re-correlating the claimants predictor variable scores with their overall weighted summative scores as described in steps 1–7 previously).

The last two columns in Table 1 display the impact or relative importance for overall fraud classifications of a high suspiciousness category "0" and a low suspiciousness category "1" answer for each variable. These columns also display the relative ranks of classification impacts for each categorical response possibility for the predictor variables on suspicion of fraud (in parenthesis). Step 8 develops an asymmetric measure of the impact an individual predictor variable response has for classification (prediction weight), a new extension.

The asymmetric rank measures are calculated as follows. For a "0" answer on variable V_t (leading to increased suspiciousness of fraud), the classification impact is $B_{t0} \times W_t^{(\infty)}$, the value of which measures the (negative score) contribution which a "0" on V_t contributes to the overall summative weighted score ordering along the latent fraud suspiciousness dimension. The rank in parentheses is the impact rank of this category of this predictor's response among other predictors' "0" answers suggestive of fraud. For example, a "0" on the variable "Vehicle not listed for private use" is indicative of fraud, and the rank shows that among all the other predictor variables whose "0" answer indicates fraud, this variable has rank 1, being the most important categorical answer option signaling fraud among all predictor variables.

The results in the last two columns of the Table are asymmetric; a "1" on a variable indicative of non-fraud is not of the same importance as a "0" on the same variable for determine fraud. For example, a "1" on "Vehicle not listed for private use," only has rank 10 among other "1's." It is not as strong a signal for non-fraud as is a "0" on this same variable for signaling fraud (rank 1 among "0" answers).

Table 1. Assessment of Spanish claimant's fraudulent property damage claim using aPRIDIT ($n = 1995$).

Fraud Predictor Variable: "0"=More Suspicion, "1" =Less Suspicion	Number of "0"	PRIDIT Score for "0" (B_{t0})	PRIDIT Score for "1" (B_{t1})	Fraud Discrimination Ability $W_t^{(\infty)}$	Relevance of Category "0" Response for Fraud Determination (Rank)	Relevance of Category "1" Response for Non-Fraud Determination (Rank)
Accident occurred in county with medium or high accident frequency (yes = 0)	979	−0.509	0.491	0.659	−0.336 (2)	0.324 (5)
Accident occurred in Northern Spain (no = 0)	1725	−0.135	0.865	0.601	−0.081 (8)	0.520 (1)
No extended 3rd party liability coverage (yes = 0)	1807	−0.094	0.906	0.486	−0.046 (10)	0.440 (2)
No police report for accident exists (yes = 0)	1773	−0.111	0.889	0.467	−0.052 (9)	0.415 (4)
Policy has no deductible (yes = 0)	1943	−0.026	0.974	0.435	−0.011 (15)	0.423 (3)
Vehicle not listed for private use (yes = 0)	232	−0.884	0.116	0.381	−0.337 (1)	0.044 (10)
Claim not reported within required time (yes = 0)	482	−0.758	0.242	0.331	−0.251 (3)	0.080 (7)
Insured accepts the blame for accident (yes = 0)	639	−0.68	0.32	0.297	−0.202 (5)	0.095 (6)
Accident occurred at night (yes = 0)	268	−0.866	0.134	0.244	−0.211 (4)	0.033 (11)
Insured has same family name as other vehicle driver, policy holder or owner (yes = 0)	125	−0.937	0.063	0.149	−0.139 (6)	0.009 (14)
Accident occurred between policy effective date and issue date (yes = 0)	33	−0.983	0.017	0.122	−0.120 (7)	0.002 (18)
No additional coverage for accessories (yes = 0)	1860	−0.068	0.932	0.052	−0.004 (17)	0.048 (9)
Witness to the accident exists (no = 0)	1981	−0.007	0.993	0.052	−0.000 (18)	0.052 (8)
Accident occurred in non-urban area (yes = 0)	143	−0.928	0.072	0.04	−0.037 (11)	0.003 (17)
Insured has record of multiple claims (yes= 0)	1225	−0.386	0.614	0.035	−0.013 (14)	0.021 (12)
Insured driver is not married (yes = 0)	682	−0.658	0.342	0.028	−0.018 (12)	0.009 (15)
Accident occurred during weekend (yes = 0)	541	−0.729	0.271	0.02	−0.014 (14)	0.005 (15)
Existence of suspicious report or unusual circumstances around the accident. (yes = 0)	1187	−0.405	0.595	0.018	−0.007 (17)	0.011 (12)

As another example, "1" on the variable "Policy has no deductible" is a much stronger indicator about the claimant *not* engaging in fraud (importance rank 3 out of 18 items) than "0" on this same variable has about the claimant engaging in fraud ("0" here rank 15 out of 18). Essentially, *having* no deductible provides little information supporting a fraud suspicion, as both fraudsters and non-fraudsters often have no deductible. However, illustrating the asymmetry in aPRIDIT, having a deductible *does* yield much more information that fraud is *not* involved since non-fraudsters are willing to commit their own money to repairs (in the form of paying a deductible) if they are involved in an accident whereas fraudsters are not. aPRIDIT identifies this asymmetric importance of the categories in predictor variables, which can be very relevant strategically and managerially.

Diagnostically, the fact that aPRIDIT provides different classification importance assessments for different categories within a predictor variable improves the social scientist's ability to know how to more accurately detect the target hidden behavior (what to focus on). In these results, some variable's answers are more important in signaling membership in the target group whose membership is to be assessed than in signaling normal behavior.

The asymmetry in aPRIDIT is important for understanding the sources of overall fraud suspicion in this application. Having knowledge of the relative importance of the components of the fraud indicators helps

decide when to contest a claim and, contra positively, knowing when there are strong indicators signaling non-fraud to help decide to settle the claim rapidly. This is practically important as hundreds of thousands of claims are handled annually. aPRIDIT can automate screening of claims so the claims adjuster can focus on the important answer categories and overall suspiciousness scores so customers *not* high on the suspiciousness dimension can be paid immediately. Thus, aPRIDIT may not only assist in identifying the hard-to-detect target behavior group of fraudsters, but also in keeping non-fraudulent customers happy.

Results such as those in Table 1 also assist the social scientist in eliminating non-discriminating predictor variables from future data collection. For example, even though the predictor variables were selected by domain experts for being indicative of fraud, some variables such as "Insured driver is not married," "Existence of suspicious report or unusual circumstances around the accident," "Accident occurred during weekend," and "Accident occurred in non-urban area" do not assist in differentiating claim suspiciousness. Elimination of these variables from future data collection saves time and money, and can assist the social science researcher in unsupervised classification by identifying unproductive classification questions or responses for elimination in future surveys. While the exact cut off in value for elimination of a predictor variable is up to the researcher, those variables with low overall variable weight $W_t^{(\infty)}$ and also high individual predictor category ranks (indicating they are not relatively very useful for discriminating no matter what response the respondent gives) are strong candidates for elimination.

Study two: identifying Portuguese high school students with heavy alcohol consumption

Teen drinking is an important issue in many countries. Heavy drinkers are more susceptible to future addiction, job and school performance deterioration, and other socially undesirable consequences. Again, heavy drinkers will not willing and truthfully disclose their alcohol abuse issues (leaving the researcher without a reliable dependent variable for detection using supervised learning methods). Schools and others can better assist if the problem is identified early. The second study applies aPRIDIT to high school alcohol abuse in Portugal, classifying high school students into heavy drinker or not heavy drinker subgroups using predictor variable indicators of higher likelihood of heavy drinking.

The Study Two dataset consists of 1044 public school student records collected by Paulo Cortez and Alice Silva from the University of Minho during 2005–2006 (Cortez & Silva, 2008). School reports and closed end questionnaires were used to construct a dataset with 32 student attributes for predicting school performance. Many of these variables were collected for other reasons and are not useful for our purposes. The reduced predictor variable set used in this study consisting of 25 dichotomous variables is presented in Table 2. The student alcohol consumption dataset was archived by Fabio Pagnotta and Hossain Mohammad Amran and is available from the University of California, Irvine Machine Learning Repository.[9] A description of the data is provided in the archive, and decisions on data transformations made are in Pagnotta and Amran (2016).

Table 1 in Pagnotta and Amran (2016) presents the predictor variables collected for heavy alcohol consumption among the high school students in the study. To be as consistent as possible with the underlying stochastic dominance assumption (monotonicity of categories along the hidden suspiciousness dimension) for aPRIDIT analysis, and using information from the published literature on drinking propensity, the binary variables' responses were ordered such that a lower score ("0") was expected to raise more suspiciousness of being a heavy drinker or was expected to be positively correlated with heavy drinking. For example, Puiu (2013) reports teen drinking is positively related to internet use, so the variable "Internet at home?" was scored with "Yes" being a "0." MacMillan (2011) reports teen dating is positively related to teen drinking so the variable "Currently in a romantic relationship?" was scored with "Yes" as "0." Teen drinking is negatively related to extracurricular activities (Joiners, 2010) so a "No" answer to "Extra-curricular activity involvement?" answered was labeled "0." In some cases (as whether the high school was Gabriel Pereira or Mousinho da Silveira), random ordering was used.

Table 2 presents the variables, the calculation of the PRIDIT scores B_{t0} and B_{t1}, the calculation of predictor variable discrimination ability measure $W_t^{(\infty)}$, and asymmetric measure of importance of response options for a heavy drinker or not group determination. Following the aPRIDIT methodology, all 1044 students were scored and ranked according to their aPRIDIT "heavy drinking suspiciousness" score, and

[9]Database available at Database: UCI machine learning repository, Student alcohol consumption data set, https://archive.ics.uci.edu/ml/datasets/Student+Alcohol+Consumption Accessed October 8, 2016.

the relative worth of each predictor variable was assessed for ability to distinguish between individuals highly suspicious of heavy drinking and those without high suspicion. Similar in structure to Table 1, Table 2 presents the analysis of predictor variables ordered by the rank importance of the predictive discrimination ability measure $W_t^{(\infty)}$.

The last two columns display the relative importance or contribution of categories "0" and "1" on that variable, respectively, for obtaining overall heavy drinker classifications. These columns also show the relative ranks or classification impacts (relevance of a "0" or a "1") for each variable on suspicion of heavy drinking (in parenthesis). These provide an asymmetric measure of the impact an individual predictor variable response has for classification (prediction weight).

The results in the last two columns of the Table 2 show asymmetry. For example, a "0" on "Past class failures," indicating that the student had failed at least one class is not very informative *in favor* of the conclusion that they are a heavy drinker (rank 12 out of 25 in terms of the contribution a "0" on this variable would impact the overall score indicative of heavy drinking) as many students, both drinkers and non-drinkers, may fail a class. On the other hand, a "1" on this same variable indicating the student has *never* failed a class is much more indicative of the student *not* being a heavy drinker (rank 3 in importance among all indicator flags for not being a heavy drinker). Similarly, while the variable "Wants to take higher education?" has the fourth largest overall predictive weight $\mathbf{W_t}^{(\infty)}$, this is due almost entirely to the fact that a "yes" on this variable is highly indicative of not being a heavy drinker (rank 1 among indicators of not being a heavy drinker). A "no" on this variable has little value for predicting heavy drinking behavior (rank 14 among categorical indicators of heavy drinking) because many students in this sample, both heavy drinker and not, do not want to take higher education.

Table 2 information also assists in eliminating non-discriminating variables from data collection. Even though the predictor variables were selected because they were expected to be correlates of heavy drinking, some variables such as "Quality of family relationships," "family size," and "Current health status" do not discriminate and are not importantly ranked for either determining if a high school student is a heavy drinker or in determining whether the high school student is not a heavy drinker. These predictor variables could be deleted. This type of information is not only useful for classification without a known dependent variable, it also can assist in streamlining questionnaire development to the most relevant questions to save time and money when studies are to be repeated over time.

Validation of the aPRIDIT methodology

While aPRIDIT shows how to create an overall "suspiciousness score" for hard-to-detect target behaviors, the issue exists as to whether this score actually relates positively and significantly to the hidden behavior whose identification is desired. Can we trust the suspiciousness of target group membership ranking produced by aPRIDIT and the classification produced by aPRIDIT?

The above question is difficult to answer in general for unsupervised classification methods (such as aPRIDIT). Validation is difficult precisely because, by definition, there is no endogenous dependent variable against which to judge the unsupervised method's classification performance. There is no straightforward way to evaluate the accuracy or validity of the algorithm absence a metric for judgement.

There is a way of validating an unsupervised classification method. If feasible, one could spend extra time or money and collect reliable dependent target group membership information on some subset of individuals. One can then judge the proposed unsupervised classification methodology by comparing the classification obtained by the unsupervised classification methodology with these known labels. This is akin to using a dataset in which the outcome is reliably known and simply not provide the aPRIDIT algorithm with this information, and then seeing how aPRIDIT performs relative to the known target membership classification.

If aPRIDIT can predict well (as measured by concordance with some trusted external grouping or classification) even without employing knowledge of a dependent variable, then it can be better trusted as a useful classification tool for decision-making in the unsupervised context where such external grouping knowledge is not available. We refer to this type of validation as "external validation" since the ability to actually gather dependent variable information for judgement purposes is external to the dataset used in unsupervised training. We now present an external validation of the aPRIDIT unsupervised classification.

Table 2. aPRIDIT assessment of Portuguese high school student's heavy alcohol consumption ($n = 1044$).

Predictor of Heavy Drinking "0" = More Likely to be a Heavy Drinker	Number of "0"	PRIDIT Score B_{t0}	PRIDIT Score B_{t1}	Predictor Discrimination Ability $W_t^{(\infty)}$	Importance of "0" for of Heavy Drinking (Rank)	Importance of "1" for Non-Heavy Drinking (Rank)
Mother's education level (0 if >9th grade)	544	−0.479	0.521	0.606	−0.316 (6)	0.290 (2)
Father's education level (0 if >9th grade)	455	−0.564	0.436	0.590	−0.257 (8)	0.333 (1)
Student's school (0 if Mousinho da Silveira)	772	−0.26	0.739	0.560	−0.414 (2)	0.146 (5)
Wants to take higher education? (0 if no)	955	−0.085	0.915	0.501	−0.458 (1)	0.043 (14)
Urban or rural student address (0 if rural)	759	−0.273	0.727	0.447	−0.325 (5)	0.122 (6)
Past class failures (0 if ≥ 1)	861	−0.175	0.825	0.435	−0.359 (3)	0.076 (12)
Internet at home? (0 if yes)	827	−0.208	0.792	0.397	−0.314 (7)	0.083 (11)
Home to school travel time (0 if ≥30 min)	943	−0.097	0.903	0.368	−0.333 (4)	0.036 (16)
Study time per week (0 if <2 h)	727	−0.304	0.696	0.353	−0.246 (9)	0.107 (8)
Paid for extra classes? (0 if no)	220	−0.789	0.211	0.337	−0.071 (17)	0.266 (3)
Family educational support? (0 if no)	640	−0.387	0.613	0.303	−0.186 (11)	0.117 (7)
Student's age (0 if ≥18)	752	−0.280	0.720	0.296	−0.213 (10)	0.083 (10)
Free time after school (0 if very high)	936	−0.103	0.897	0.183	−0.164 (12)	0.019 (23)
Attended nursery school? (0 if no)	835	−0.200	0.800	0.176	−0.141 (14)	0.035 (17)
Extra educational support? (0 if no)	119	−0.886	0.114	0.172	−0.020 (21)	0.153 (4)
Frequency of going out with friends (0 if very high)	881	−0.156	0.844	0.170	−0.144 (13)	0.027 (19)
Extra-curricular activity involvement? (0 if no)	516	−0.506	0.494	0.164	−0.081 (15)	0.083 (9)
Currently in a romantic relationship? (0 if yes)	673	−0.355	0.645	0.115	−0.074 (16)	0.041 (15)
Student's guardian (0 if not mother)	728	−0.303	0.697	0.096	−0.067 (19)	0.029 (18)
Quality of family relationships (0 if ≤2 on 5 point very bad to very good)	967	−0.074	0.926	0.075	−0.070 (18)	0.006 (24)
Parent's cohabitation status (0 if living apart)	121	−0.884	0.116	0.063	−0.007 (24)	0.056 (13)
Sex (0 if female)	591	−0.434	0.566	0.050	−0.028 (20)	0.022 (21)
Number of absences (0 if >5)	317	−0.696	0.304	0.032	−0.010 (21)	0.023 (20)
Family size? (0 if ≤3)	738	−0.71	0.29	0.029	0.021 (23)	0.009 (22)
Current health status (0 if very good)	395	−0.622	0.378	0.001	0.000 (23)	0.001 (25)

aPRIDIT external validation results using known group designations in study one

The Spanish fraud data involve identifying the hard-to-detect behavior of insurance consumers' defrauding an insurance company, a behavior consumers actively try to conceal. Using the aPRIDIT classification methodology we obtain an overall suspicion score for each consumer filing an insurance claim. For this study a claimant was labeled as belonging to the target "fraud group" if their overall aPRIDIT suspicion score $S^{(\infty)}$ was negative and non-fraudulent if their score was positive.

We can externally validate aPRIDIT here because the insurance company took the extra time and effort to obtain "true" fraud classification label for a subset of consumer claims. They were able to provide this "dependent variable" information to us since when the Spanish insurer had significant suspicion of fraud being perpetrated the insurer denied complete payment or announced to the consumer that their insurance contract would be canceled. Negotiations with the customer ensued and, consequently, the fraudulent customer could (and often did) admit fraud. As a result of this admission, no further punitive action was taken (other than canceling the claim without penalty).

When a customer admits fraud, the insurer has no uncertainty that fraud existed. Thus, for these there is information concerning fraud group membership on a

Table 3. Established fraud and non-fraud cases versus aPRIDIT fraud prediction for the Spanish dataset; 67% over-all agreement, 74% agreement on actual target group (fraud) cases.

Cross-classification Results for aPRIDIT and Fraud

	Established non-fraud	Established fraud	aPRIDIT totals
aPRIDIT Classified as Non-Fraud	593	257	850
aPRIDIT Classified as Fraud	405	740	1145
Actual Total	998	997	1995

Sensitivity = 740/997 = 0.742, Specificity = 593/998 = 0.594.

subgroup. Because legal prosecution of customers for insurance fraud is extremely rare in Spain, such fraud admissions are more common in Spain than in USA. We externally validate aPRIDIT's classification performance by comparing aPRIDIT classification to classification elicited by the insurer from the claimant.

The Spanish insurance company provided data on customers' claims with half (998) being labeled by the insurer as legitimate claims and half (997) identified as fraudulent claims[10]. In this Table 3, a claimant was classified by aPRIDIT as fraudulent if their overall summative PRIDIT score was negative. Table 3 compares the results of aPRIDIT versus the known fraud classification for the Spanish dataset.

Table 3 shows a reasonably high degree of classification concordance result between aPRIDIT and the insurance company's actual fraud determination ((593 + 740)/1995 = 67% as per Table 3). This provides support for aPRIDIT's external validity. Moreover, for the hard-to-detect subgroup of fraudsters (the focus of actionable fraud analysis), aPRIDIT did correctly identify 74% (=740/997) of the known fraud cases as likely fraud. It should also be noted that company determination of fraud/non-fraud classification is costly in terms of time and human capital, whereas aPRIDIT utilizes the common directionality in the construction of predictor variable construction (based on a priori expert judgment, with lower categorical responses on predictor variables expected to be more indicative of "suspicious of fraud") in a methodology that can be automated.

Another important goal of insurance fraud claim detection is to ascertain which, among hundreds of thousands of claims, should be paid immediately and which are suspicious enough to warrant transferring to a special investigative fraud unit. Note that aPRIDIT performed well "blindfolded" (without having target group labels) and can be automated to produce a suspiciousness ranking technique based on the internal consistency of ordering of the categorical responses across the ensemble of predictor variables. Strategically, the ability to linearly order all claim files along the hidden "suspicious of fraud" dimension allows the researcher (or claim manager in this case) to focus attention on the claims most likely to be fraudulent. The need for efficient and effective (rapid) suspicion category classification research is, in part, because there are regulatory penalties in many countries (e.g. USA) for delayed payment of valid insurance claims. Since external validation shows aPRIDIT produces an ordinal overall claim suspiciousness score highly effective in identifying the target group members, the insurer can order claims so the most suspicious can be investigated first. The high degree of concordance with the known group memberships labels when they exist allows increased confidence in aPRIDIT use in situations where an actual dependent variable does not exist (i.e. the typical unsupervised training case).

Comparison of aPRIDIT with other unsupervised classification methods

There are a few competing quantitative unsupervised analysis methods available for detecting hidden behaviors or making a classification determination when the delineating dependent variable cannot be reliably assessed (i.e. due to social undesirability bias). Next, we discuss some alternatives.

Comparison with Kohonen Self-Organizing Feature Maps

One unsupervised method is Kohonen's Self-Organizing Feature Map, a neural network method (Kohonen, 1989). This has been used in automobile fraud detection (cf., Brockett, Derrig, & Xia, 1998). There are several advantages of aPRIDIT over this methodology. First, aPRIDIT produces an overall suspiciousness score allowing rank ordering on the underlying latent suspiciousness of hard-to-detect target group membership, whereas Kohonen maps do not provide metric level information. Ranking on a suspiciousness dimension is managerially relevant for strategic decision-making. Additionally, an individual's aPRIDIT weighted overall score is a metric level score that can be incorporated as variable input in

[10]This oversamples of the fraud subgroup (estimates of consumer fraud in this industry are about 10% rather than 50%).

conjunction with other variables, such as demographics, if subsequent investigative analysis is desired. According to Davison, Davenport, Chang, Vue, and Su (2015), the usefulness of a scoring methodology (such as aPRIDIT provides for each individual) is increased if the scores can be used for predicting an external criterion. By contrast, Kohonen classification is graphical in nature (Brockett et al., 1998), and would be hard to incorporate into other numerical statistical analysis. In addition, Kohonen maps do not provide unique classification information whereas aPRIDIT does.

Comparison with cluster analysis

Another often-used unsupervised method for classification analysis is cluster analysis. In identifying hidden or hard-to-detect behavior, however, cluster analysis suffers because derived clusters may not have direct interpretations in terms of the specific target group behavior that is desired to be identified. Even if there emerges (or we force) two groups/clusters, we are not certain *a priori* which group to label as belonging to the target group and which to the non-target group without any additional outside information. Moreover, individual entities within these two designated clusters are not themselves metrically ordered according to suspiciousness level, so subsequent quantitative examination is not facilitated.

We did perform a cluster analysis on the Spanish consumer fraud data forcing two clusters (consistent with the fraud/non-fraud dichotomy). To resolve the intrinsic uncertainty in cluster analysis as to which cluster gets the "Fraud" label, we took the most favorable view of the clustering algorithm and designated as the fraud cluster that which had the highest overlap with the company's fraud determination of claimants. We found that aPRIDIT significantly outperformed cluster analysis when judged according to the true classification data supplied by the Spanish insurer.

Comparison of supervised logistic regression with unsupervised aPRIDIT classification

The external validation investigation exploited having "true" classifications for a subset of data. When a data set is developed for which known target group membership labels are available, then supervised classification techniques are possible. Hence, we also judged the classification strength of unsupervised aPRIDIT with logistic regression (LOGIT). We compared to Logit since, in a performance comparison of several popular supervised classification methods by Viaene, Derrig, Baesens, and Dedene (2002), logistic regression was found to perform about as well (or better) than competitors with easier computations, interpretation and communication characteristics. Viaene et al. (2002) recommend this technique, even over neural networks. Moreover, logistic regression is a familiar benchmark commonly used in social science applications. Bulut, Davison, & Rodriguez (2017), for example, use logistic regression classification accuracy when assessing subscore reliability in the profile analysis they investigated. We do not expect aPRIDIT to classify as well as supervised training methods since these use more information (the knowledge of the dependent variable) and optimize classification performance in a parametric discrimination setting, whereas aPRIDIT does not use this information and is non-parametric. As expected, supervised logistic regression classified slightly better *overall* than unsupervised aPRIDIT (overall correct classification of 73% for Logit vs. 67% for aPRIDIT). Interestingly, when focused on the important subclass identified by the insurer as fraudulent consumers, a different picture emerged. Within this subclass, aPRIDIT had a 74% agreement with the insurer's fraud class whereas Logit had a lower 70% agreement. For the target behavior group of interest (fraud), aPRIDIT outperformed Logit. In fact, within the target fraud group, 98% of the claims correctly identified by logistic regression were also correctly identified by aPRIDIT without utilizing supervision. As per West, Brockett, and Golden (1997), if there are strong differential costs between committing a Type 1 versus Type 2 classification error, improved performance in identifying members of the more important fraud group can lead to significant monetary savings with little loss of performance.

Summary and conclusions

Social scientists have long been aware that individuals can exhibit hard-to-identify behavior, and can engage in dishonest or illegal behavior that they will not self-disclose. This paper addresses an important and neglected issue concerning identification of such hidden behavior: How to identify *which* individuals are members of the target group who exhibit the behavior and are failing to honestly reveal such information. The methodological difficulty arises because there is no labeled set of individuals with membership along with known predictive covariates that can be used to "train" supervised learning techniques like logistic

regression, discriminant analysis, or probit analysis. The capability to classify hidden behaviors without access to a dependent membership variable is increasingly important.

Methodologically, detecting individuals engaging in socially undesirable or stigmatizing behavior that they will not self-disclose is more difficult than identifying "normal" behavior or motivations/behaviors that will be revealed. The dishonest person has no interest in truthfully revealing their behavior so there is no available dependent variable for training standard statistical models. This calls for an unsupervised classification methodology that is not relying on a dependent variable available.

This article introduces PRIDIT analysis (Principal Component Analysis of RIDITs) to the social science literature, and presents an asymmetric extension (aPRIDIT) that provides further directional information on the value of individual predictor variables' categorical response options for assessing membership in a dichotomous population where some people have the target behavior to be identified and some do not. We assume there are ordinal categorical predictor variables available related to the target behavior desired to be identified. aPRIDIT is a new unsupervised learning statistical classification technique that allows researchers to more sensitively classify people into groups in order to identify potential people when getting a training sample for supervised detection is not feasible.

After presenting the underlying methodological algorithm, this research illustrates aPRIDIT for identifying hidden group membership in two different international social science contexts. Using Spanish insurance company claims data, aPRIDIT identifies individuals committing insurance claim fraud based on gathered data that raises suspicion of fraud, but without access to whether or not fraud actually occurred. Using data on Portuguese high school students, aPRIDIT identifies individuals that have a high likelihood of being heavy alcohol users.

In external validation investigations, aPRIDIT performed well compared to known group memberships. The aPRIDIT technique was externally validated by acquiring a data set that actually does have the dependent variable for target group membership. In this external validation assessment, aPRIDIT does very well, and does especially well on the identification of the target behavior group members.

The proficiency of aPRIDIT in grouping individuals without prior classification information and with less restrictive statistical assumptions than supervised learning techniques such as Logit is shown to be another characteristic of aPRIDIT. Without using a training sample, aPRIDIT is able to classify performance comparable to Logit that utilized a training sample, and it even outperformed Logit performance in identifying individuals in the important target behavior group. Moreover, among true target behavior group members, 98% of those individuals identified by Logit were also identified by aPRIDIT, and without the use of any class membership knowledge. If the goal is to identify members of the target behavior group, then there may be little justification for expending the time, effort, and cost of obtaining a class membership label needed for supervised learning techniques. aPRIDIT may suffice at lower cost.

Since aPRIDIT is an unsupervised classification technique new to the literature, it was also compared to other known unsupervised classification methods such as Kohonen's self-organizing feature maps and cluster analysis. aPRIDIT outperforms both techniques in terms of the classification performance, and also in terms of the type of information produced (e.g. aPRIDIT provides a metric level numerical "suspiciousness of target behavior group membership" score capable of being used in further analysis along with other data such as demographics, etc.). Cluster analysis and Kohonen maps do not give the additional metric level "suspiciousness of target group membership" information on individuals capable of linearly ordering individuals according to suspiciousness that aPRIDIT does, and neither gives information at the individual categorical response level concerning variable classification importance.

The asymmetric "ranking of the usefulness for classification" metric provided by the aPRIDIT technique is an important extension to PRIDIT for assessing the impact of predictor variable individual categories, and for understanding the sources of heightened target group membership likelihood. aPRIDIT answers the question of why a predictor variable is important for classification, and what parts of the predictor variable's response options are most important for prediction of target behavior group membership.

There are many opportunities for future research using aPRIDIT such as for jury selection. Potential jurors may be unwilling to report with accuracy their willingness to convict, but aPRIDIT's classification expertise using individual attitudes may help determine more accurately how potential jurors might decide. The same thing is true for other societal issues and concerns, such as sexual misconduct, willingness to vote for a minority or an extremist candidate, or

any other areas where people might not be accurately disclosing. Of course, as mentioned several times before, the careful choice of predictor variables that are appropriate, consistently ordered, and all of which have reasonable association expected with the unknown latent "suspiciousness of target group membership" latent variable is of utmost importance for the aPRIDIT methodology to perform well. Classifications by aPRIDIT may be more accurate than self-reports and can be automated to help human decision-making.

Finally, aPRIDIT can assist in classification surveys or questionnaire design by signaling what questions and even what particular answer categories are most useful for gathering the desired hidden target group membership information. The asymmetric nature of aPRIDIT allows both the identification of which predictor variables have the most information for the classification, and which particular specific answers to those questions are most impactful in decision-making. Again, being able to focus on the most relevant questions for the survey's information goals can help the social science researcher save both time and money by reducing questionnaire length while still obtaining the classification desired for individuals, and a metric level "suspiciousness score" capable of input into subsequent analysis.

Article information

Conflict of interest disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: The work of Drs. Patrick Brockett and Linda Golden was supported by Grant SAB2003-0191 from the Spanish Ministry of Science, Grant 2004PIV1-00009 from Generalitat de Catalunya, Spain, and the Riskcenter, Department of Econometrics, University of Barcelona, Spain. The work of Dr. Montserrat Guillen was supported by Grant ECO2016-76203-C2- 2-P from ICREA Academia and the Spanish Ministry of Science. The work of Dr. Danae Manika was supported by a grant from the Center for Risk Management at the University of Texas at Austin, USA.

Role of the funders/sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Acknowledgments: The authors would like to thank the anonymous reviewers and also the Editor Peter C.M. Molenaar and the Associate Editor Keith Widaman for their comments on prior versions of this manuscript which significantly improved the paper. The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors' institutions is not intended and should not be inferred.

References

Ai, J., Brockett, P., Golden, L. L., & Guillen, M. (2013). A robust unsupervised method for fraud rate estimation. *Journal of Risk and Insurance*, *80*, 121–1431. doi:10.1111/j.1539-6975.2012.01467.x

Artís, M., Ayuso, M., & Guillén, M. (1999). Modeling different types of automobile insurance fraud behavior in the Spanish market. *Insurance: Mathematics and Economics*, *24*, 67–81. doi:10.1016/S0167-6687(98)00038-9

Artís, M., Ayuso, M., & Guillén, M. (2002). Detection of automobile insurance fraud with discrete choice models and misclassified claims. *Journal of Risk & Insurance*, *69*, 325–340. doi:10.1111/1539-6975.00022

Association of Certified Fraud Examiners (2019, January 10). *2018 global study on occupational fraud and abuse* [PDF]. Retrieved from https://www.acfe.com/uploadedFiles/ACFE_Website/Content/rttn/2018/RTTN-Government-Edition.pdf

Bechtel, G. G. (1985). Generalizing the Rasch model for consumer rating scales. *Marketing Science*, *4*, 62–73. doi:10.1287/mksc.4.1.62

Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science*, *17*, 235–255. doi:10.1214/ss/1042727940

Brockett, P. L. (1981). A note on the numerical assignment of scores to ranked categorical data. *The Journal of Mathematical Sociology*, *8*, 91–101. doi:10.1080/0022250X.1981.9989917

Brockett, P. L., Derrig, R. A., Golden, L. L., Levine, A., & Alpert, M. (2002). Fraud classification using principal component analysis of RIDITs. *Journal of Risk & Insurance*, *69*, 341–372. doi:10.1111/1539-6975.00027

Brockett, P. L., Derrig, R. A., & Xia, X. (1998). Using Kohonen's self-organizing feature map to uncover automobile bodily injury claims fraud. *The Journal of Risk and Insurance*, *65*, 245–274. doi:10.2307/253535

Brockett, P. L., & Golden, L. L. (1992). A comment on "Using rank values as an interval scale" by Dowling and Midgley. *Psychology and Marketing*, *9*, 255–261. doi:10.1002/mar.4220090307

Brockett, P. L., & Levine, A. (1977). On a characterization of RIDITs. *The Annals of Statistics*, *5*, 1245–1248. doi:10.1214/aos/1176344010

Bross, I. D. H. (1958). How to use RIDIT analysis. *Biometrics*, *14*, 18–38. doi:10.2307/2527727

Bulut, O., Davison, M. L., & Rodriguez, M. C. (2017). Estimating between-person and within-person subscore reliability with profile analysis. *Multivariate Behavioral Research*, *52*, 86–104. doi:10.1080/00273171.2016.1253452

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, *41*, 1. doi:10.1145/1541880.1541882

Cortez, P., & Silva, A. (2008). Using data mining to predict secondary school student performance. In A. Brito & J. Teixeira (Eds.). *Proceedings of 5th future business technology conference (FUBUTEC 2008)* (pp. 5–12). Porto, Portugal, April 2008, Eurosis.

Daultry, S. (1976). Principal components analysis. *Concepts and techniques in modern geography*, No. 8 (pp. 1–51). Norwich: Geo Abstracts Ltd., University of East Anglia..

Davison, M. L., Davenport, E. C., Jr., Chang, Y.-F., Vue, K., & Su, S. (2015). Criterion-related validity: Assessing the value of subscores. *Journal of Educational Measurement*, *52*, 263–279. doi:10.1111/jedm.12081

De Jong, M. G., Pieters, R., & Fox, J. P. (2010). Reducing social desirability bias via item randomized response: An application to measure underreported desires. *Journal of Marketing Research*, *47*, 14–27. doi:10.1509/jmkr.47.1.14

Gates, K. M., Lane, S. T., Varangis, E., Giovanello, K., & Guiskewicz, K. (2017). Unsupervised classification during time-series model building. *Multivariate Behavioral Research*, *52*, 129–148. doi:10.1080/00273171.2016.1256187

Gliem, J. A., & Gliem, R. R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. *2003 Midwest research to practice conference in adult, continuing, and community education* (pp. 82–88). Columbus, OH. Retrieved from http://hdl.handle.net/1805/344

Golden, L. L., & Brockett, P. L. (1987). The effects of alternative scoring techniques on the analysis of rank ordered categorical data. *The Journal of Mathematical Sociology*, *12*, 383–414. doi:10.1080/0022250X.1987.9990021

Golub, G., & Van Loan, C. F. (1996). *Matrix computations* (3rd ed.). Baltimore, MD, USA: The Johns Hopkins University Press.

Joiners, F. (2010). Teen drug use and extracurricular activities. *Alcohol self-help news*. Retrieved from https://alcoholselfhelpnews.wordpress.com/2010/09/13/teen-drug-use-and-extracurricular-activities

Kohonen, T. (1989). *Self-organizing feature maps. Self-organizing and associative memory* (3rd ed.). Berlin, Heidelberg: Springer-Verlag.

MacMillan, A. (2011, September 28). Teen dating may spread teen drinking. Retrieved April 21, 2019, from http://www.cnn.com/2011/09/28/health/teen-dating-drinking-relationship

Pagnotta, F., & Amran, H. M. (2016). *Using data mining to predict secondary school student alcohol consumption*. Camerino, Italy: University of Camerino. doi:10.13140/RG.2.1.1465.8328

Puiu, T. (2013). Teen drinking linked to internet use. Retrieved April 21, 2019, from https://www.zmescience.com/research/teenage-drinking-link-internet-434324/

Viaene, S., Derrig, R. A., Baesens, B., & Dedene, G. (2002). A comparison of state-of-the-art classification techniques for expert automobile insurance claim fraud detection. *Journal of Risk & Insurance*, *69*, 373–421. doi:10.1111/1539-6975.00023

West, P., Brockett, P. L., & Golden, L. L. (1997). A comparative analysis of neural network and statistical methods for predicting consumer choice. *Marketing Science*, *16*, 370–391. doi:10.1287/mksc.16.4.370

Woodside, A. G. (2013). Moving beyond multiple regression analysis to algorithms: Calling for adoption of a paradigm shift from symmetric to asymmetric thinking in data analysis and crafting theory. *Journal of Business Research*, *66*, 463–472. doi:10.1016/j.jbusres.2012.12.021